UBER – Assignment Case Study
Demand & Supply Gap Analysis

Anugraha Sinha

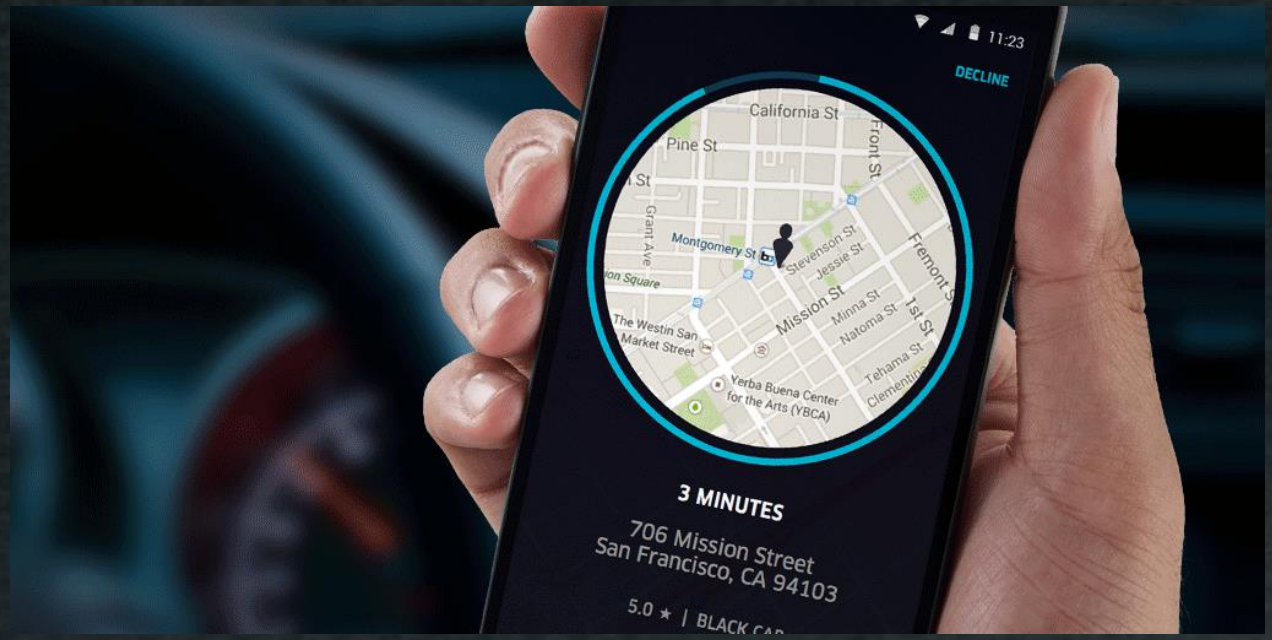
Roll No: DDA1710381

UBER

Analysis Objectives

❑Business Objective

- Identify root cause for the following problem(s):
  - Cancellation of Cabs
  - Non-Availability of Cabs
- Possible hypothesis of the problem(s)
- Recommend solution for driving the changes to meet the challenges.



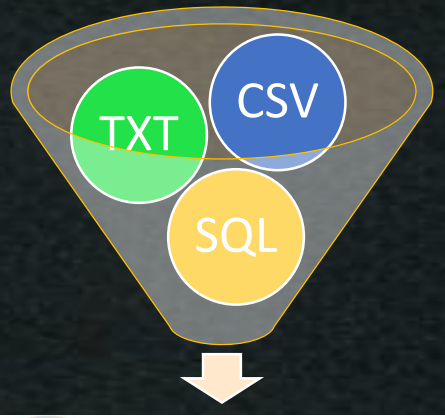
❑Metadata

- Trips to-and-from the City & Airport
- Trip Status - “Completed/Cancelled/No Cars”
- Data of 5 days (weekdays)
- Spanning complete 24 hours
- ~6700+ trip information

Important Note :

In this presentation, 2 analysis for trips originating at City is shared. The 2nd analysis for City is presented in the end of the presentation after demand/supply based analysis. Evaluation committee is requested to take a look at the last analysis also.

Methodology



DATA SOURCING

Data Cleaning

Data Understanding

Business Driven Derived Metrics

Data Driven Derived Metrics

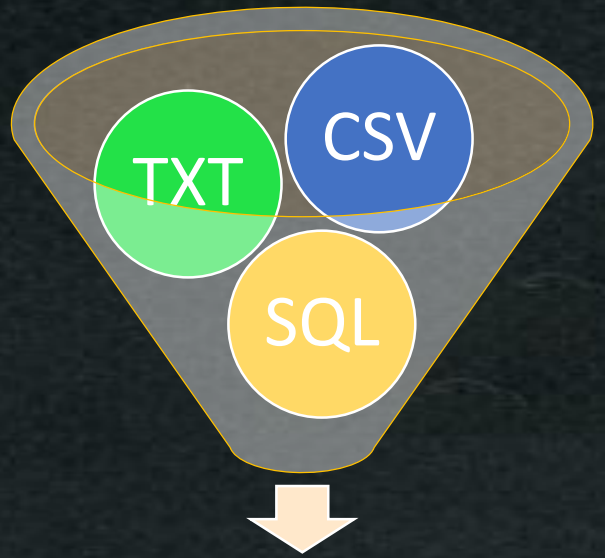
Univariate Analysis

Segmented Univariate Analysis

Bivariate Analysis

Data Sourcing – Cleaning (Insights)

- CSV based data
  - 6745 Rows, 6 features
  - 4 numeric features & 2 character features



Missing Value/NA/Blank Analysis			
Column Name	**No. of NA(s)**	**No. of Blanks**	**Reasons (After investigation)**
All Data Points	6564	None	Total of Driver.id + Drop.timestamp
Driver.id	2650	None	Number of NAs for status = “No Cars” -> 2650. (Same as Driver.id = NAs)
Drop.timestamp	3914	None	Number of NAs for “Cancelled/No Cars” trips = 3914 (Same as drop.timestamp = NAs)
TOTAL	2650+ 3914 = 6564	Final Conclusion	No changes to existing data set, NAs pegged to assist analysis later on

Note:
All graphs plotted in this presentation are marked with G(XX) value. A similar variable can be looked for as “GXX” in the R file for the code to generate the graph.

Glimpse of Standardized DataSet

With new Business/Type Driven Derived Metrics

Request.id	Pickup.point	Driver.id	Status	Request.timestamp	Drop.timestamp	Request.timestamp_2	Drop.timestamp_2	Request.date	Request.hour	Drop.date	Drop.
619	Airport	1	Trip Completed	11/7/2016 11:51	11/7/2016 13:00	2016-07-11 11:51:00	2016-07-11 13:00:00	2016-07-11	11	2016-07-11	
867	Airport	1	Trip Completed	11/7/2016 17:57	11/7/2016 18:47	2016-07-11 17:57:00	2016-07-11 18:47:00	2016-07-11	17	2016-07-11	
1807	City	1	Trip Completed	12/7/2016 9:17	12/7/2016 9:58	2016-07-12 09:17:00	2016-07-12 09:58:00	2016-07-12	9	2016-07-12	
2532	Airport	1	Trip Completed	12/7/2016 21:08	12/7/2016 22:03	2016-07-12 21:08:00	2016-07-12 22:03:00	2016-07-12	21	2016-07-12	
3112	City	1	Trip Completed	13-07-2016 08:33	13-07-2016 09:25	2016-07-13 08:33:00	2016-07-13 09:25:00	2016-07-13	8	2016-07-13	
3879	Airport	1	Trip Completed	13-07-2016 21:57	13-07-2016 22:28	2016-07-13 21:57:00	2016-07-13 22:28:00	2016-07-13	21	2016-07-13	
4270	Airport	1	Trip Completed	14-07-2016 06:15	14-07-2016 07:13	2016-07-14 06:15:00	2016-07-14 07:13:00	2016-07-14	6	2016-07-14	
5510	Airport	1	Trip Completed	15-07-2016 05:11	15-07-2016 06:07	2016-07-15 05:11:00	2016-07-15 06:07:00	2016-07-15	5	2016-07-15	
6248	City	1	Trip Completed	15-07-2016 17:57	15-07-2016 18:50	2016-07-15 17:57:00	2016-07-15 18:50:00	2016-07-15	17	2016-07-15	
267	City	2	Trip Completed	11/7/2016 6:46	11/7/2016 7:25	2016-07-11 06:46:00	2016-07-11 07:25:00	2016-07-11	6	2016-07-11	
1467	Airport	2	Trip Completed	12/7/2016 5:08	12/7/2016 6:02	2016-07-12 05:08:00	2016-07-12 06:02:00	2016-07-12	5	2016-07-12	

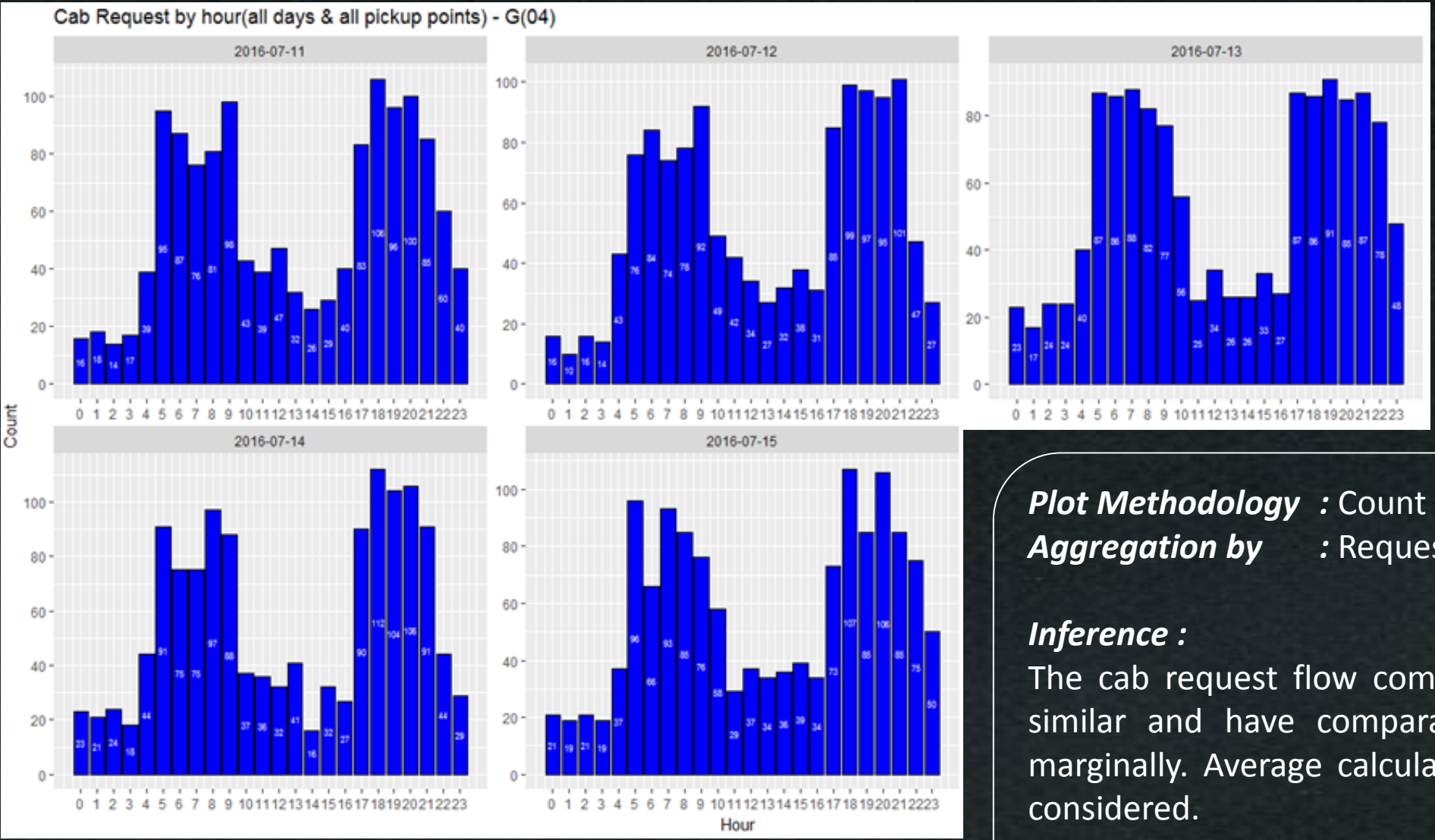
Date/time provided are not in consistent format.

Date/time standardized for uniformity

Type of Derived Metrics	Base Columns	New Columns
Type Driven Derived Metrics	Request.timestamp_2	Request.date,Request.hour
Type Driven Derived Metrics	Drop.timestamp_2	Drop.date,Drop.hour
Business Driven Derived Metrics	Drop.timestamp_2,Request.timestamp_2	Trip.length

Data Driven Derived Metrics

Divide hours of day under segments



Count of Cab Requests across all hours of the day & across all days.

Plot Methodology **:** Count
Aggregation by **:** Request.date & Request.hour

Inference :

The cab request flow compared over the days looks similar and have comparable metrics, which differ marginally. Average calculation over the days can be considered.

Data Driven Derived Metrics

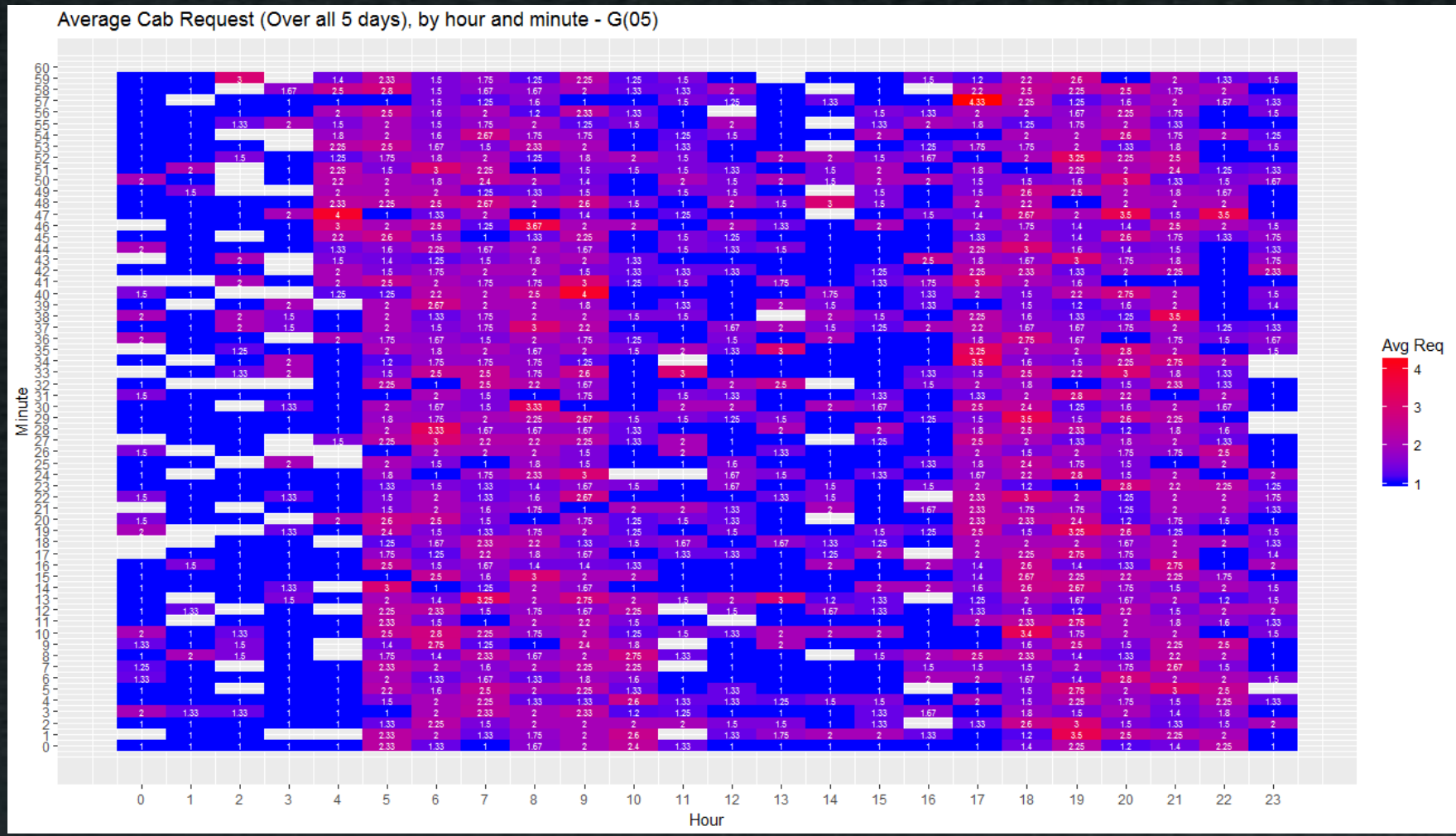
Divide hours of day under segments (Contd.)

In continuation with Slide 6 inference, distribution of cab requests on a granular hours & minute basis

Plot Methodology : Average
Aggregation by : Request.date,
: Request.hour
: Request.min

Inference :
The whole day can be divided in 5 different segments as per below timings

Time of the Day	Day Segment Name
0000 hrs – 0359 hrs	1_LateNight
0400 hrs – 0559 hrs	2_EarlyMorning
0600 hrs – 1059 hrs	3_MorningRush
1100 hrs – 1659 hrs	4_Day
1700 hrs – 2359 hrs	5_EveningRush



Univariate Analysis

Inference :

G(01)

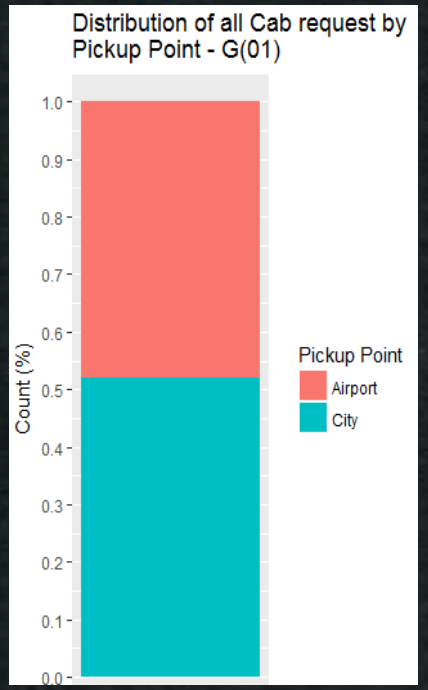
The masked data set given has been evenly distributed between pickup.points of airport and city.

G(02)

%Count of “No Cars” + “Cancelled” is higher than “Trip Completed”. This problem will be addressed in this analysis.

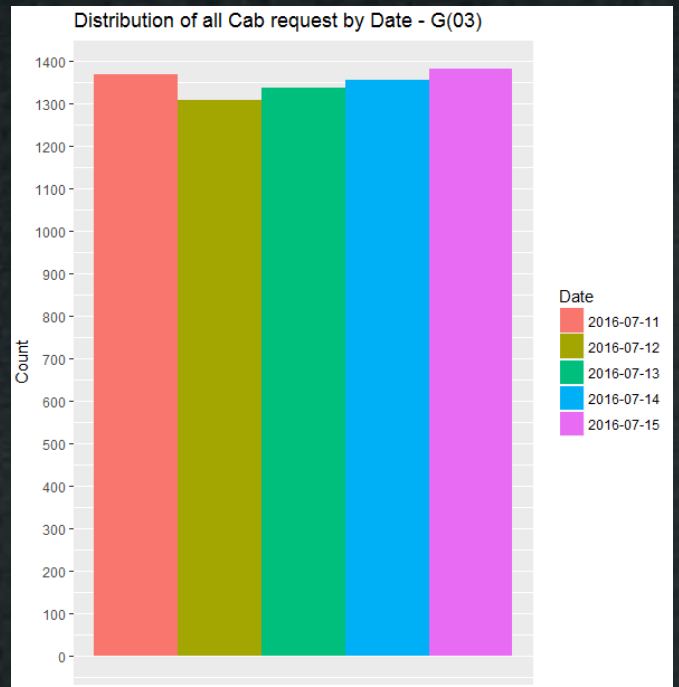
G(03)

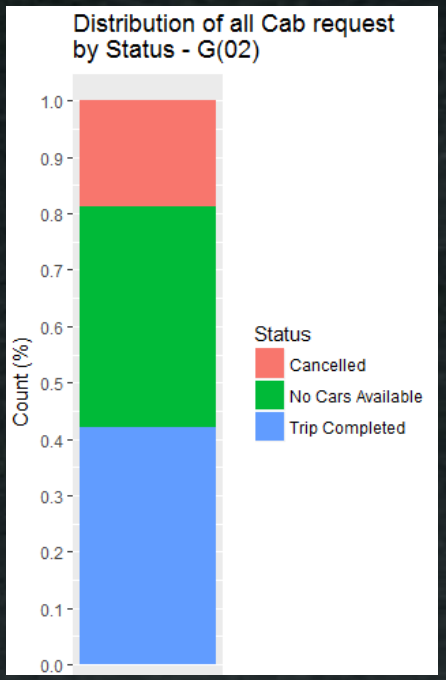
The masked data set given has comparable data points for all days (~1350-1380 per day)



Plot Methodology : Count

Aggregation : Various (as per plot)

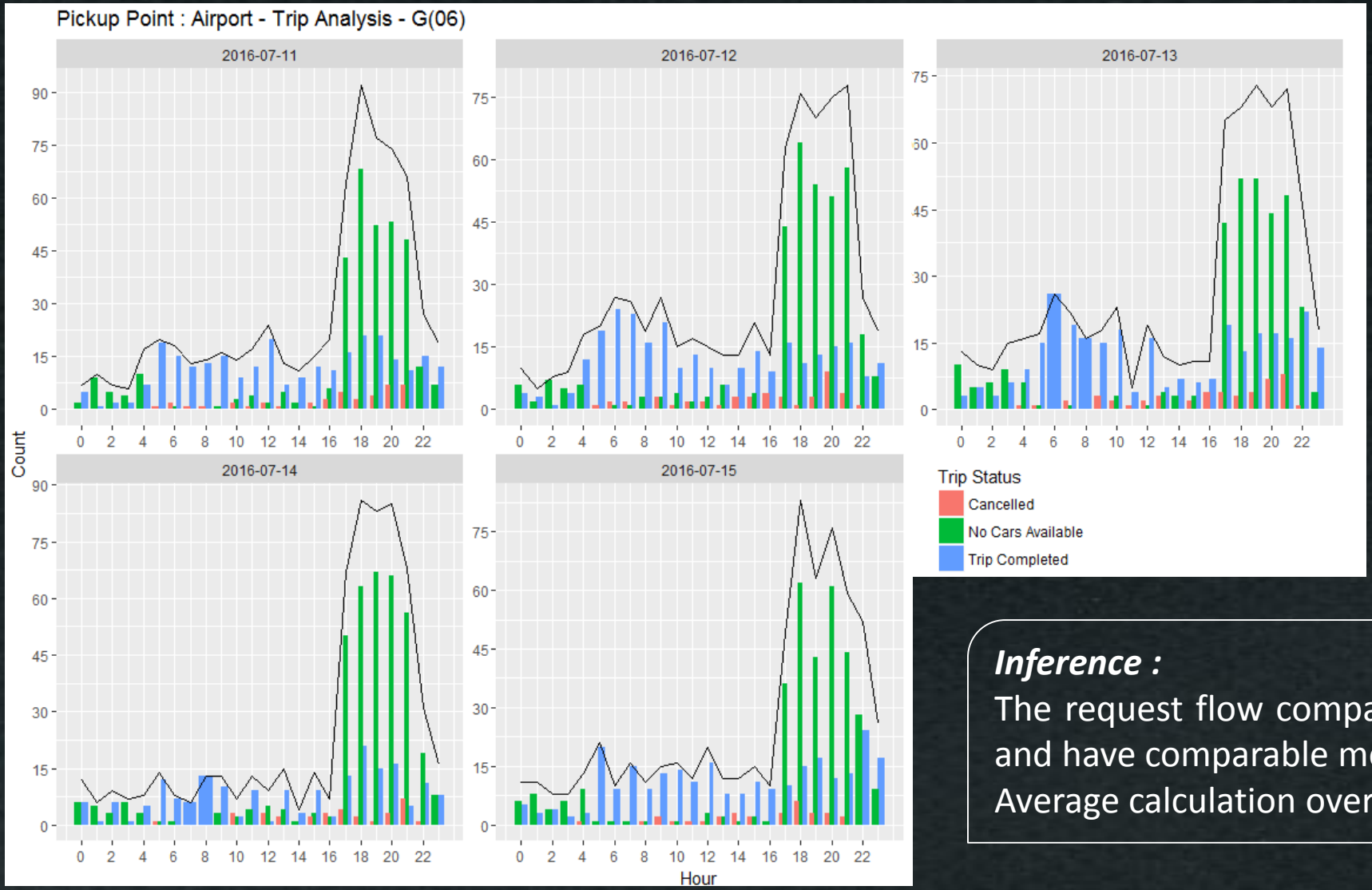




Logical Assumptions Taken For Analysis

- Cancellation are being done by drivers & not by riders (There is no method to differentiate this from the data set given, so assuming drivers are cancelling request)
- Definition of Demand (at any pickup.point or time of the day)
  - Trip Completed + Cancelled + No Cabs Available
- Definition of Supply
  - Trip Completed + Cancelled
- Cancelled trips are considered within “Supply” definition because even if the ride got cancelled by driver, but organic supply for the cabs does exist. Reasons for such cancellation need to be analyzed separately.
- The timezone is IST.

Analysis for Trips starting @ Airport

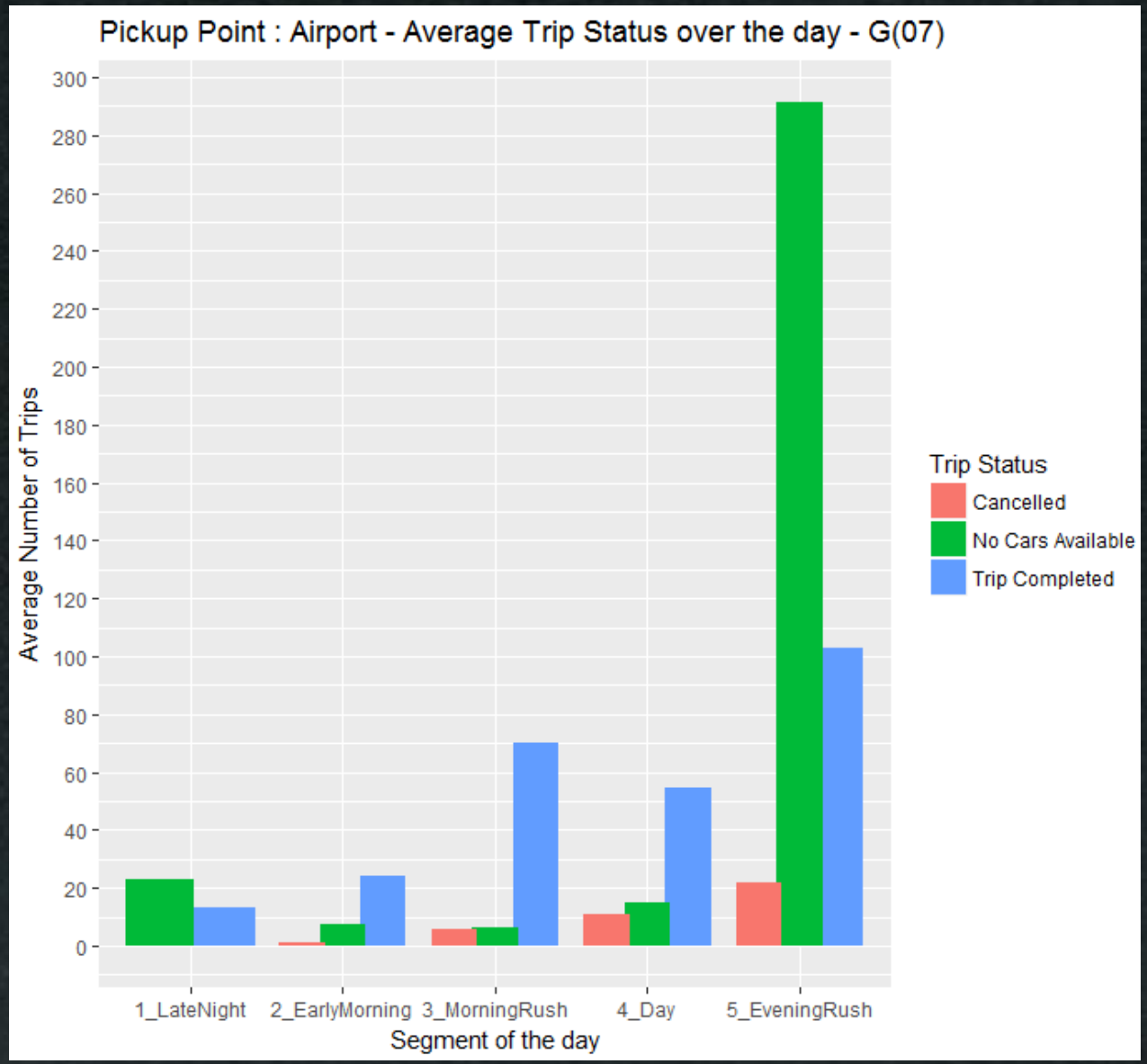


Count of Cab Requests from City -> Airport across all hours of the day & across all days.

Plot Methodology : Count
Filter : Pickup = City
Aggregation by : Request.date & Request.hour

Inference :
The request flow compared over the days looks similar and have comparable metrics, which differ marginally. Average calculation over the days can be considered.

Analysis for Trips starting @ Airport (Contd.)



Average Count of Cab Requests from Airport -> City Averaged over all the days, and clubbed on the basis of Day segments

Plot Methodology	*:* Average
Filter	*:* PickupPoint = Airport
Aggregation by	*:* Request.date & Day Segment

Inference :

EveningRush hours have strange spike in number of "No Cars Available" problem. The organic supply (Trips Completed + Cancelled), is way less than the total demand.

Average Demand : 410 ~ 420 (Definition as per Slide 9)

Average Supply : 120 ~ 125 (Definition as per Slide 9)

Demand and Supply gap should be checked for all days to verify if similar pattern is seen.

Analysis for Trips start @ Airport
Demand Supply Gap

Demand & Supply difference graph during EveningRush hours based on categorization of Trip Status (NoCars/Cancelled/Completed) as per bins of Demand/Supply for trips from Airport

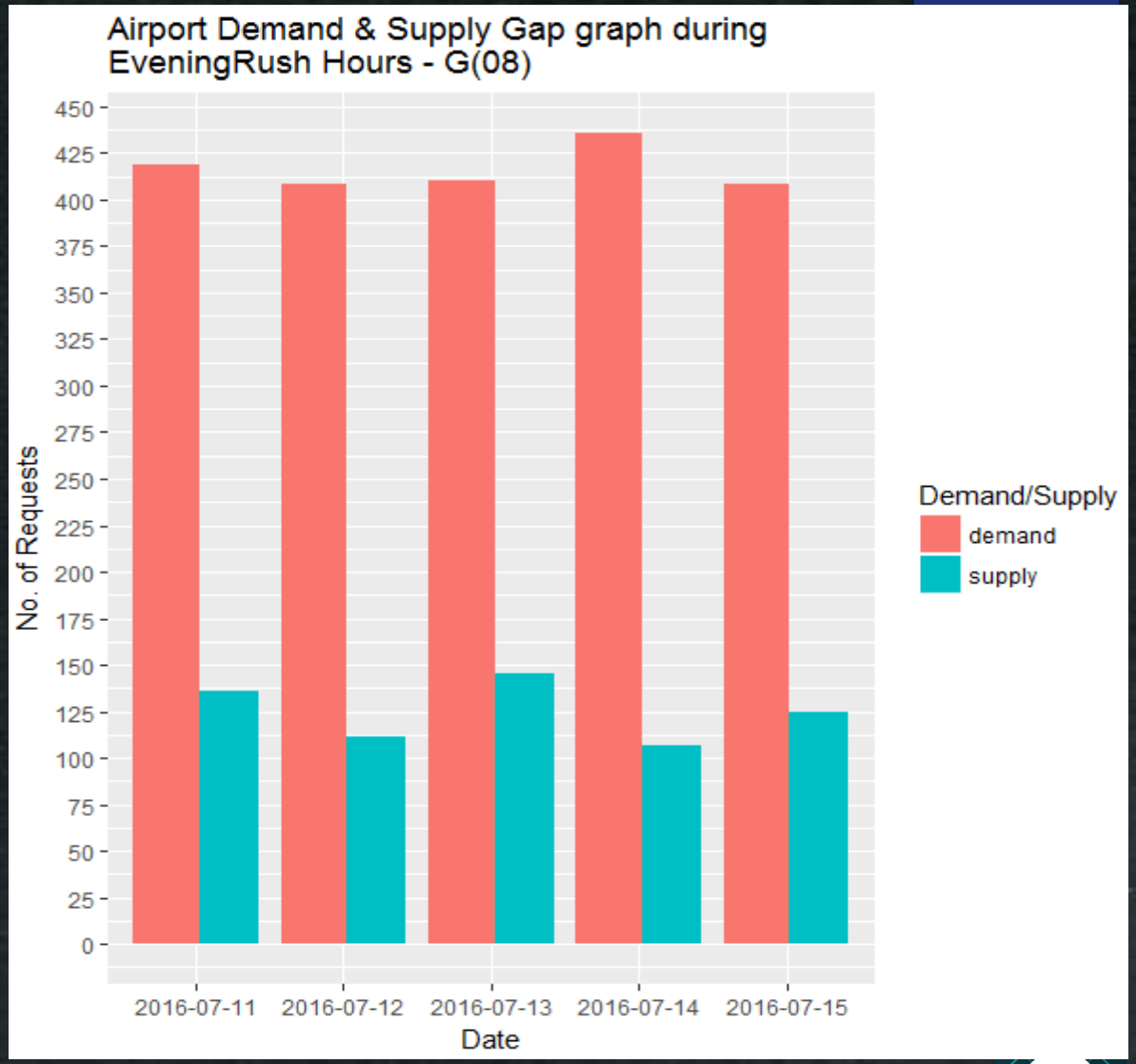
Plot Methodology	*:* Count
Filter	*:* PickupPoint = Airport
	: DaySegment = "EveningRush"
Mutated Column	*:* Demand/Supply
Aggregation by	*:* Request.date & Demand.Supply

Inference :

Across all days for EveningRush hours for trip starting from the Airport, there a huge demand & supply gap.

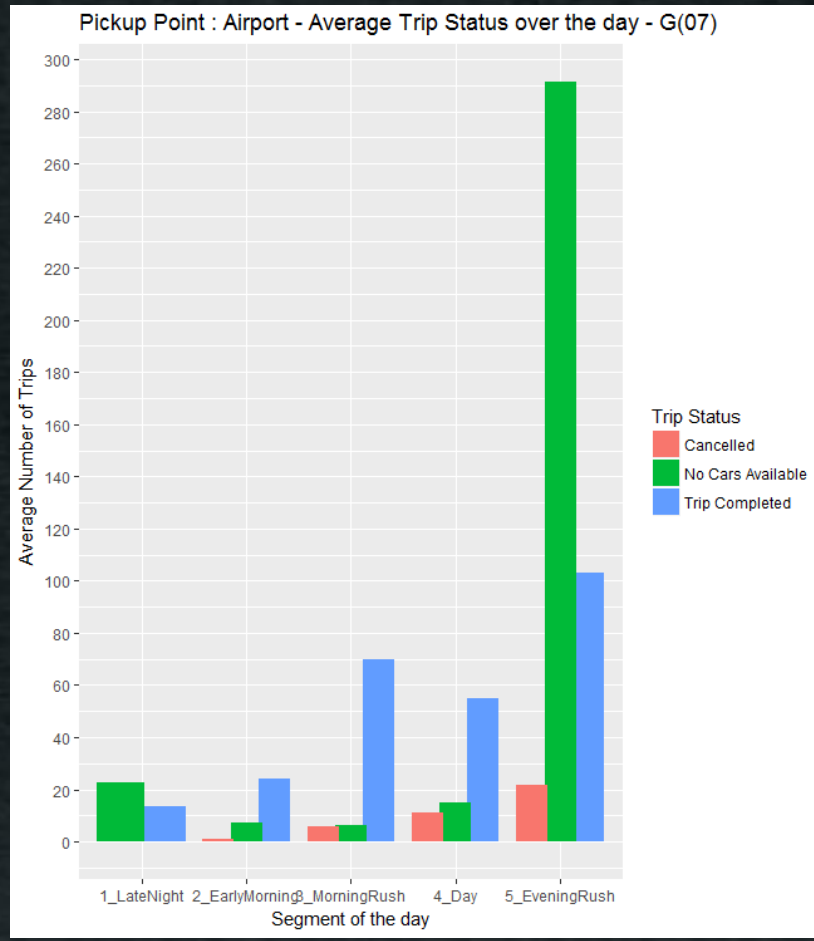
The supply only suffices 25%-35% of what the demand is.

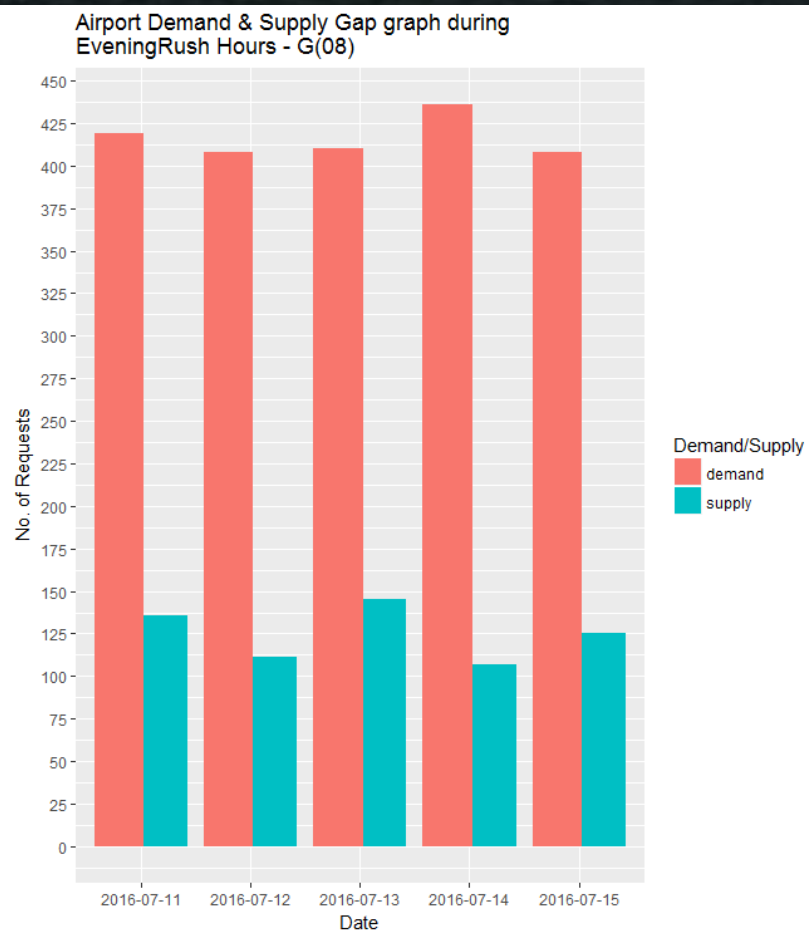
	Request.date	demand	supply	Ratio(Supply/Demand)
1	2016-07-11	419	136	0.3245823
2	2016-07-12	408	111	0.2720588
3	2016-07-13	410	145	0.3536585
4	2016-07-14	436	107	0.2454128
5	2016-07-15	408	125	0.3063725



Analysis for Trips start @ Airport Recommendations

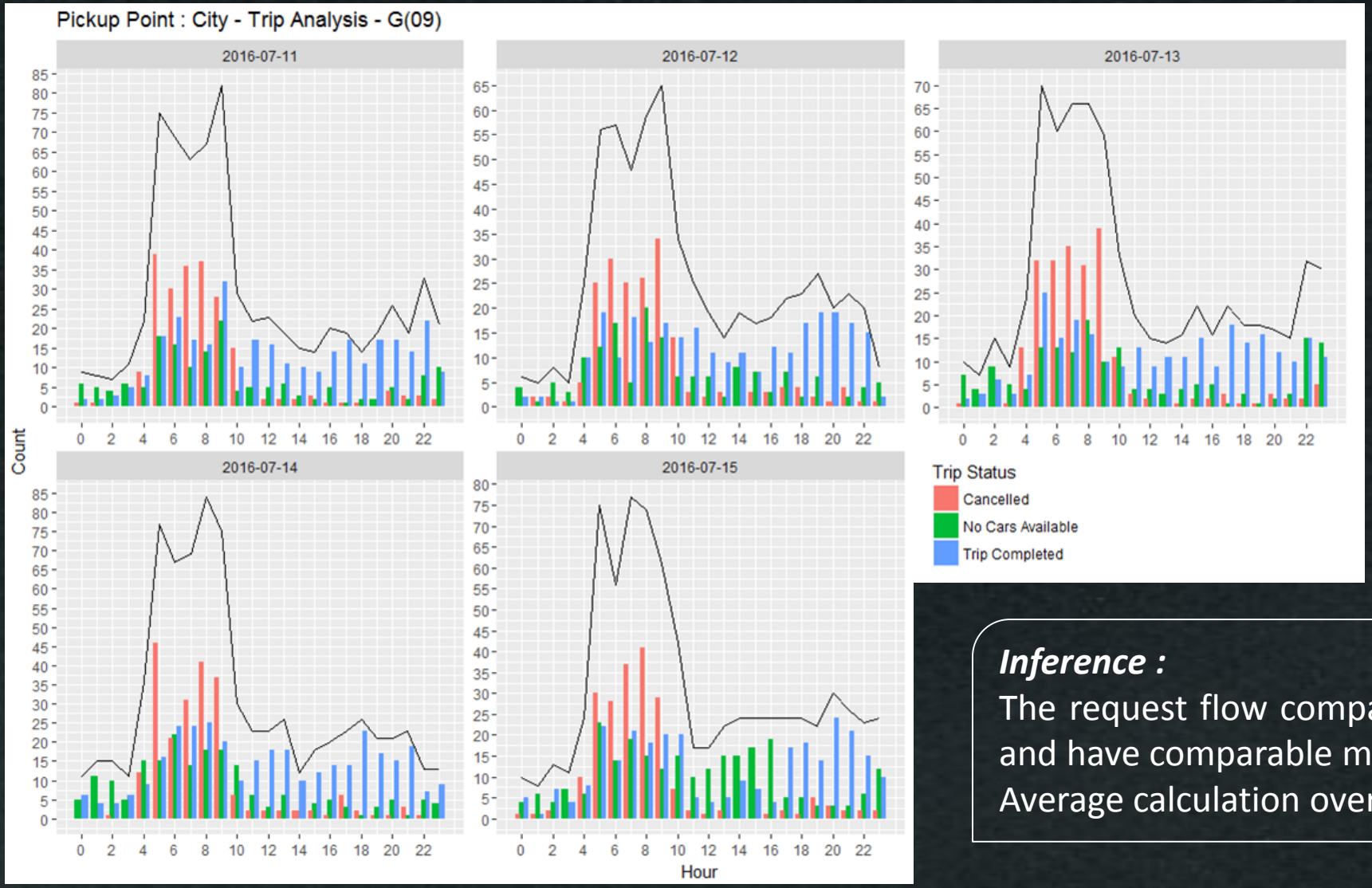
- ***Final Conclusion***
  - For EveningRush hours, the volume of traffic is very high.
  - This is potential revenue augmenting time for drivers and Uber
  - There is a strong gap between the demand and supply during this period that needs to be bridged
- ***Recommendations***
  - Extra incentives (greater percentage of fare share) can be released to drivers for trips taken up from Airport in the evening.
  - Uber network may start prioritizing trips to near-by areas of airport starting from 1530-1600hrs, so that supply increases in evening
  - A tie up with AAI (Airport Authority of India) for its employees for commuting in the evening may increases supply at airport in the evening.





- Driver association with Uber network needs to be strengthened further. Ground teams need to cover last-mile associates and members to bring more drivers on the road.

Analysis for Trips starting @ City

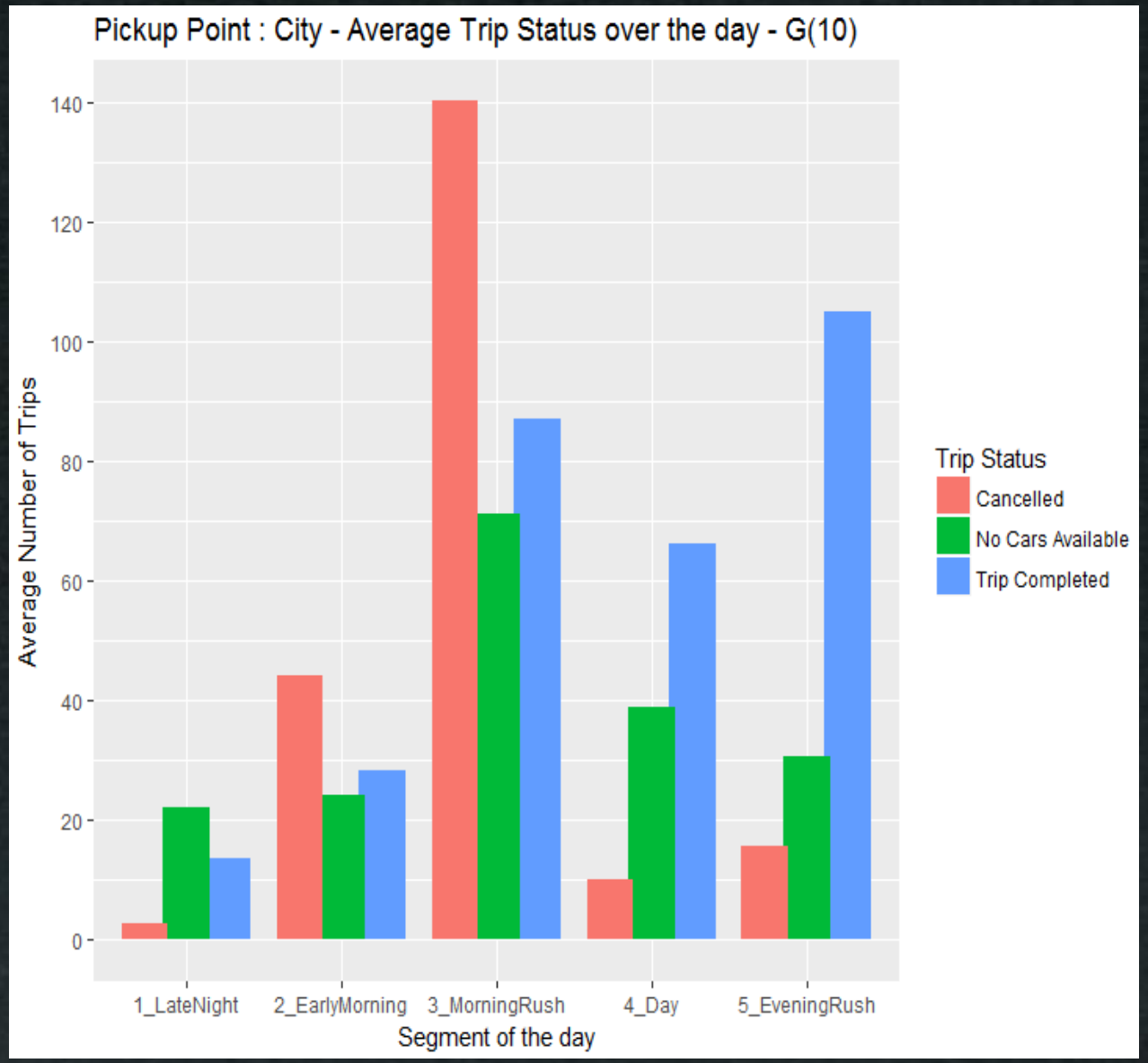


Count of Cab Requests from Airport -> City across all hours of the day & across all days.

Plot Methodology : Count
Filter : Pickup = Airport
Aggregation by : Request.date & Request.hour

Inference :
The request flow compared over the days looks similar and have comparable metrics, which differ marginally. Average calculation over the days can be considered.

Analysis for Trips starting @ City (Contd.)



Average Count of Cab Requests from City -> Airport Averaged over all the days, and segmented on the basis of Day segments trimmed as per previous slides

Plot Methodology	*:* Average
Filter	*:* PickupPoint = City
Aggregation by	*:* Request.date & Day Segment

Inference :

EarlyMorning & MorningRush hours have high number of "Cancellations" & "No Cars". The organic supply (Trips Completed + Cancelled), is way less than the total demand.

Average Demand (EarlyMorning) : ~ 100
Average Supply (EarlyMorning) : ~ 70

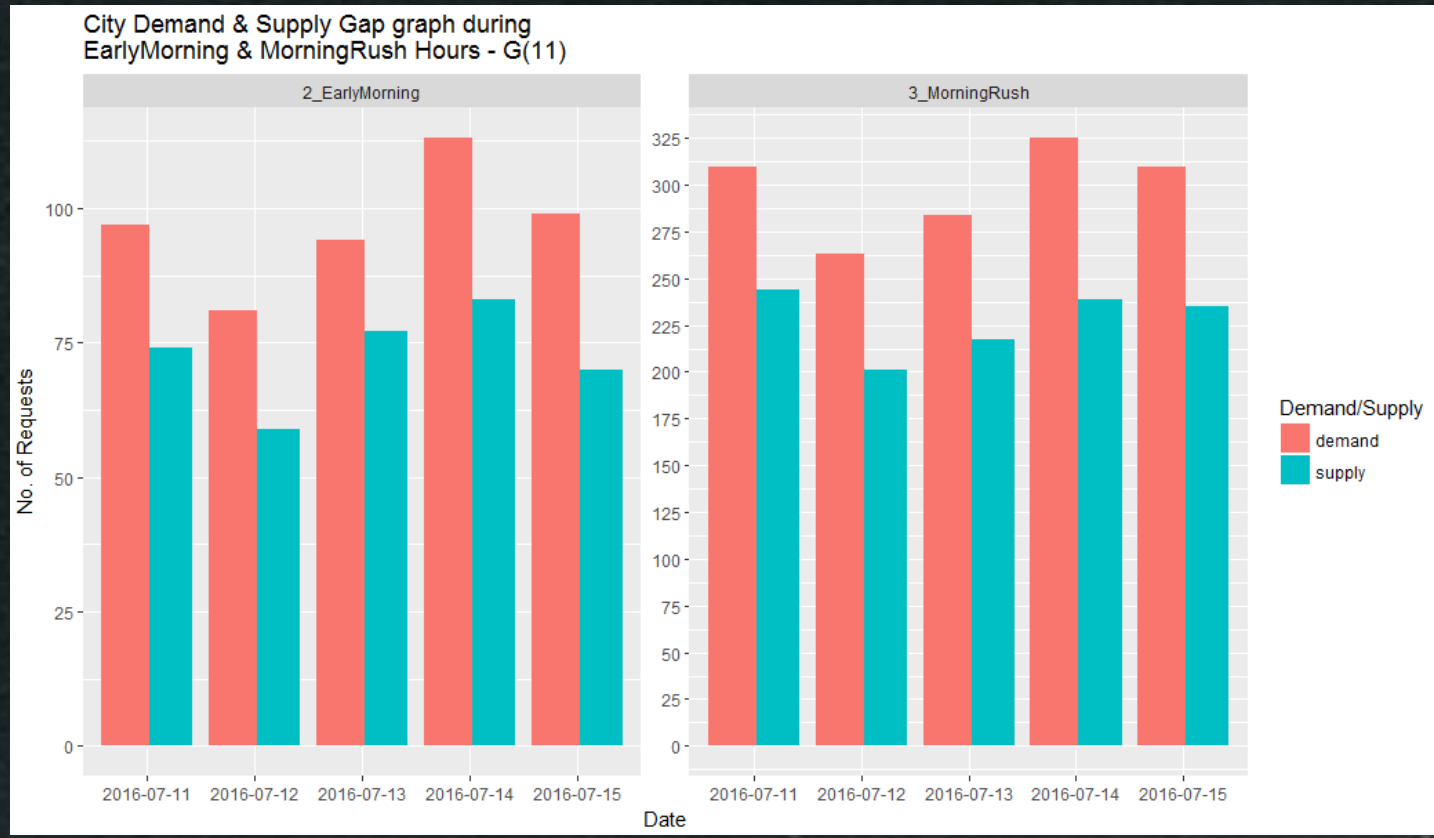
Average Demand (MorningRush) : ~ 300
Average Supply (EarlyMorning) : ~ 160

Demand and Supply gap should be checked for all days

Analysis for Trips starting @ City
Demand Supply Gap

Demand & Supply difference graph during EarlyMorning & Morning Rush hours based on categorization of Trip Status (NoCars/Cancelled/Completed) as per bins of Demand/Supply for trips from City

Plot Methodology : Count
Filter : PickupPoint = City
: DaySegment = “EarlyMorning” &
: DaySegment = “MorningRush”
Mutated Column : Demand/Supply
Aggregation by : Request.date & Demand.Supply



Inference :
For trips starting at City, the demand supply gap is clear from the graphs across all days. The supply is seems to be falling short by ~ 30%.

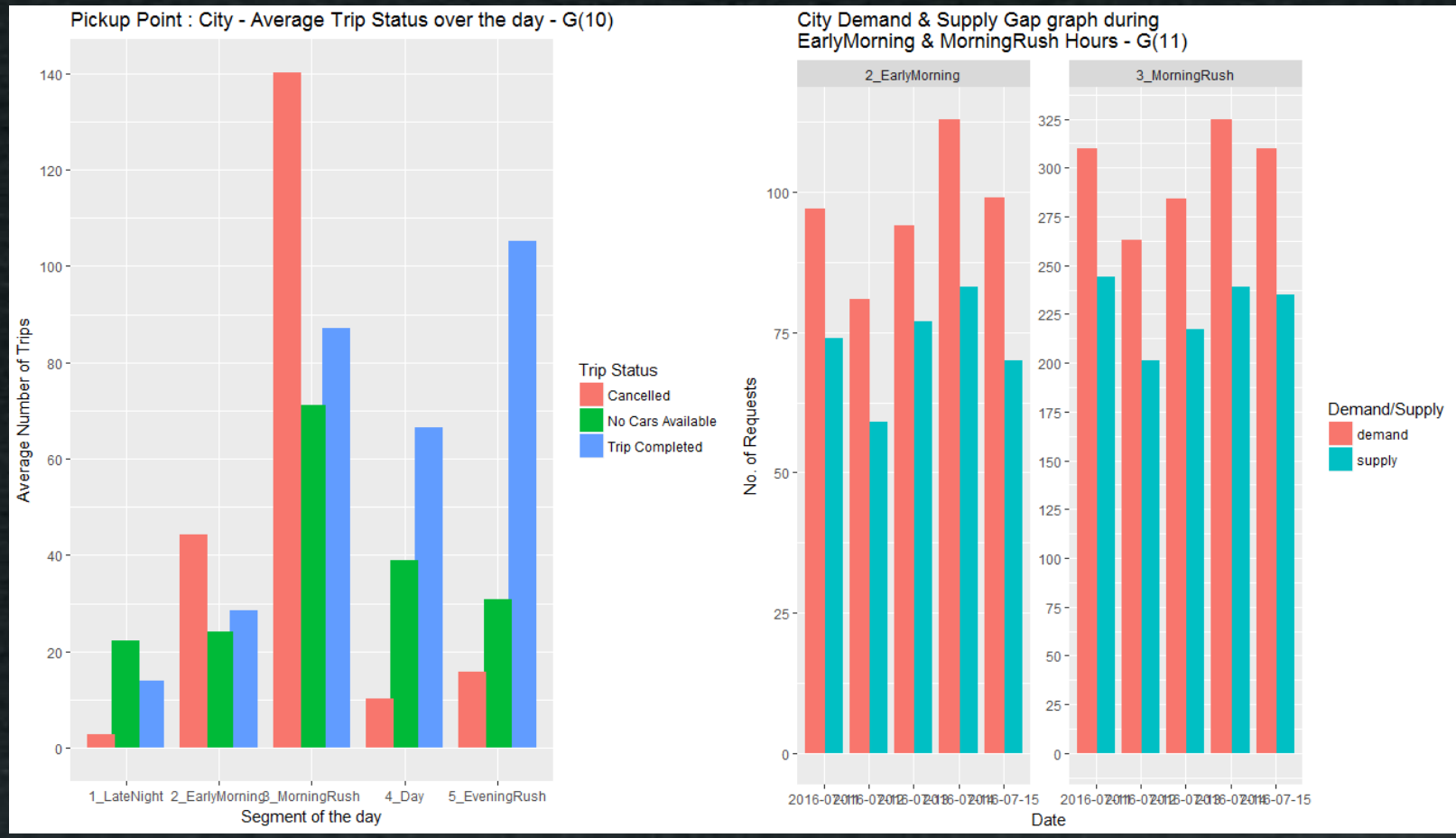
Analysis for Trips starting @ City Recommendations

❑ *Final Conclusion*

- For EarlyMorning and MorningRush hours, the volume of traffic is highest in city when compared with other times of the day.
- The augmented traffic volume is not being converted to “Trip Completed” leading to loss of potential revenue.
- The reason is an evident demand supply gap that needs to be bridged.

❑ *Recommendations*

- Uber field operation teams should get in touch with individual drivers to understand the reason for cancellation in the morning hours. (Another hypothesis is also presented after this slide)
- Uber may plan to concentrate certain marketing initiatives in the areas near the airport, so that drivers could be shown the incentives of taking trips to the airport.



- A tie up with AAI (Airport Authority of India) for its employees getting free from night-shifts will increase round-trip flow to-and-from the airport.

Analysis for Trips starting @ City Based on Waiting Time at Airport

Logical Assumptions Taken For Analysis (In addition to the ones mentioned in Slide 9)

- For Every driver id, a set of 2 trips are selected which match following criteria
  - 1st Trip Completed from City -> Airport
  - 2nd Trip (May or may be completed) from Airport -> City
  - Both 1st Trip (City – Airport) and 2nd Trip Request (Airport – City) should happen on the same Request.date

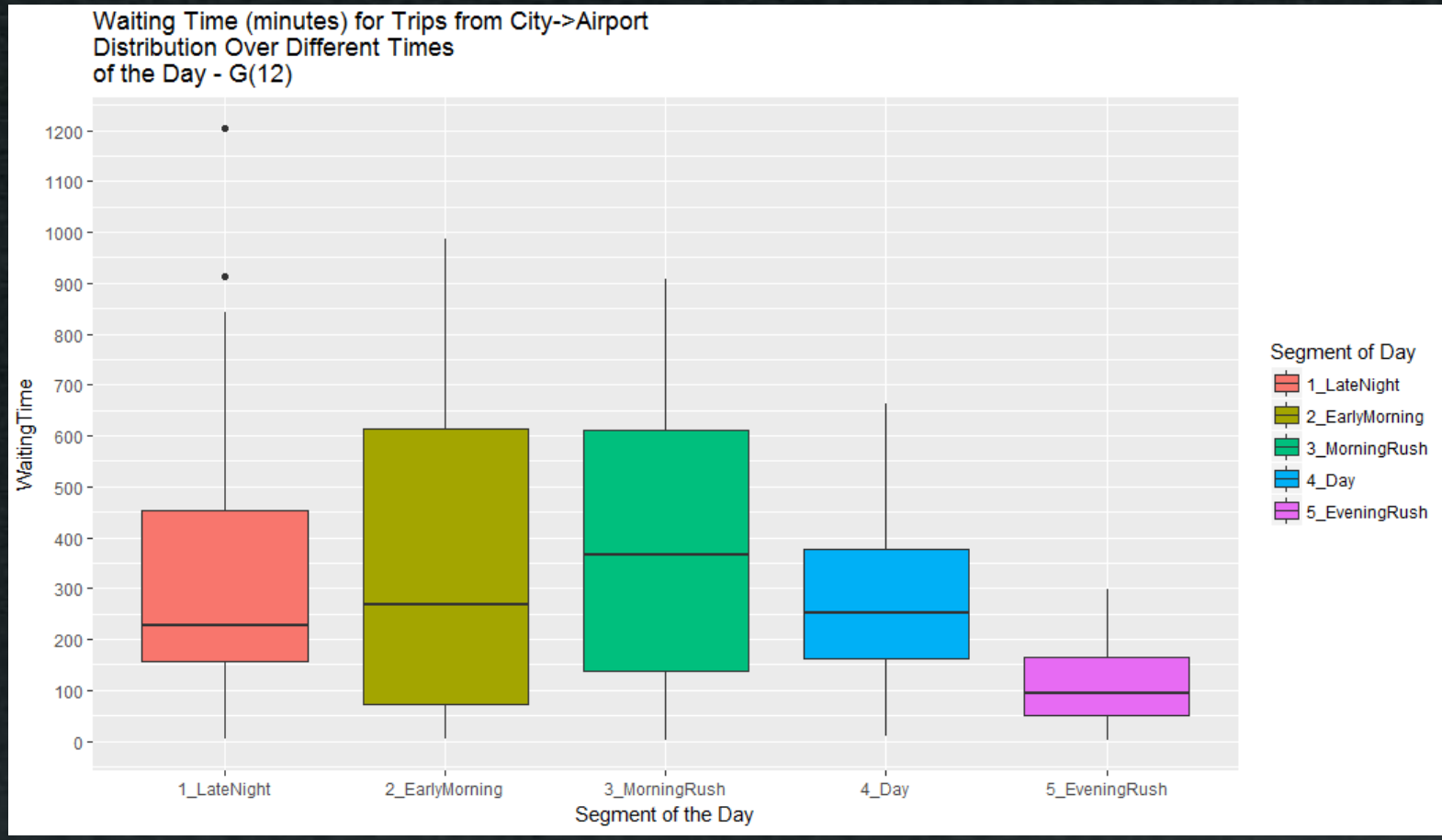
Request.id	Pickup.point	Driver.id	Status	Request.timestamp	Drop.timestamp	Request.timestamp_2	Drop.timestamp_2	Request.date	Request.hour	Drop.date
1807	City	1	Trip Completed	12/7/2016 9:17	12/7/2016 9:58	2016-07-12 09:17:00	2016-07-12 09:58:00	2016-07-12	9	2016-07-12
2532	Airport	1	Trip Completed	12/7/2016 21:08	12/7/2016 22:03	2016-07-12 21:08:00	2016-07-12 22:03:00	2016-07-12	21	2016-07-12
4652	City	2	Trip Completed	14-07-2016 12:01	14-07-2016 12:36	2016-07-14 12:01:00	2016-07-14 12:36:00	2016-07-14	12	2016-07-14
5023	Airport	2	Cancelled	14-07-2016 19:04	NA:NA	2016-07-14 19:04:00	NA	2016-07-14	19	NA

- Based on the above segregation criteria, following data was obtained
  - Total rows in given data = 6547 rows
  - Selected wait time rows = 810 rows
- In view of the above it can be said that 1 out of 8 rows from the complete data set have been taken for wait time calculation. Given the constrained of the given data set, 1/8 is a considerably good sample to reflect on the population for waiting time.
- Assuming that masked data set is for drivers that ply only between airport & city. (Since we do not have complete data from Uber for all kinds of trips, for real life analysis, this assumption becomes necessary)
- Such waiting time was analyzed for trip from city -> airport.

Analysis for Trips starting @ City Based on Waiting Time at Airport

❑ *Final Conclusion*

- For the trips taken in the “EarlyMorning” & “MorningRush” hours, the waiting time is very high.
- Waiting time reduces as the day progresses and is minimum in the evening. Which is evident from Airport analysis that due to large inflow of flights in the evening, demand increased at airport in the evening.



❑ *Recommendations*

- Extra incentives (greater percentage of fare share) can be released to drivers for trips taken up from City during morning hours.
- A tie up with AAI (Airport Authority of India) for its employees working in night hours for getting back to city will reduce the wait time at the airport.
- The demand supply gap analysis does exist along with this problem. However, it seems this problem is more pressing and in accordance with driver’s psyche when compared with previous analysis.

Final Result of the Analysis
Q3 & Q4

- *Supply & Demand Gap*
  - There is clear supply and demand gap for both the pickup points
  - Airport have more severe demand supply problem as compared to city.
- *Reasons*
  - One of the primary reasons is the lack of adequate organic supply. Along with this, high wait time for drivers at the airport is discouraging drivers to take up trips from the city in the morning hours. Additionally, during evening hours there is huge gap between the demand and supply at the airport.
- *Recommendation*
  - More cabs/drivers need to get associated to meet the demand/supply gap. Field operation teams need to reach out to last-mile associates and members to strengthen the association with Uber network.
  - For reducing wait time, Uber should concentrate marketing initiatives at areas near the airport to reduce wait time for drivers at airport and keeps more cabs mobile and reduce stationary time.
  - Partnership/association with Airport Authority of India, for its employee transfer program will stimulate 24X7 traffic movement to & from the airport, as airports function round the clock.
  - Extra incentives (greater percentage of fare share) can be released to drivers for trips taken up from Airport in the evening.